Physica A 505 (2018) 500–511

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Predicting epidemic threshold of correlated networks: A comparison of methods

Xuan-Hao Chen [a,b], Shi-Min Cai [a,b,c,*], Wei Wang [b,d], Ming Tang [b,e], H. Eugene Stanley [c]

[a] Web Sciences Center, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China
[b] Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 610054, China
[c] Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA
[d] Cybersecurity Research Institute, Sichuan University, Chengdu 610065, China
[e] School of Information Science Technology, East China Normal University, Shanghai 200241, China

HIGHLIGHTS

- Precisely predicting outbreak threshold of correlated networks is of great Importance.
- Four theoretical methods are comparative to predicting outbreak threshold.
- The outbreak threshold in SIR model is affected by topological structure of correlated networks.
- The simulations based on SIR model evaluate the fitness of four theoretical methods.
- The CM and DMP methods perform better to predicting outbreak threshold of correlated networks.

ARTICLE INFO

Article history:
Received 25 December 2017
Received in revised form 2 March 2018
Available online 7 April 2018

Keywords:
Epidemic threshold
Theoretical methods
SIR spreading dynamics
Correlated network

ABSTRACT

Being able to theoretically predict the outbreak threshold of an epidemic is essential in disease control. Real-world correlated networks are ubiquitous and their topological structure strongly affects the prediction accuracy of present-day theoretical methods. Quantifying their accuracy and fitness in predicting outbreak thresholds of correlated networks is thus essential. We use a susceptible–infected–removed (SIR) model to examine four widely-used theoretical methods – the heterogeneous mean-field (*HMF*), quenched mean-field (*QMF*), dynamical message passing (*DMP*), and connectivity matrix (*CM*) methods – to predict the outbreak threshold of a correlated network. The potential topological structure of correlated network impacts on prediction accuracy of these four methods. We emphasize that the quantitative changes of degree correlation, degree distribution exponent and network size strongly affect the outbreak of SIR spreading dynamics, and compare the simulation results with the theoretical ones obtained from these four methods. The extensive experiments in synthetic networks show that (a) the increasing degree correlation coefficients reduce outbreak threshold and suppress outbreak size; (b) the increasing degree distribution exponents raise outbreak threshold but suppress outbreak size; (c) the increasing network sizes decrease outbreak threshold but do not affect outbreak size; (d) as for four theoretical methods, *CM* and *DMP* are more likely to precisely predict outbreak threshold because they to some extent incorporate network topology with dynamical

* Corresponding author at: Web Sciences Center, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China.
E-mail address: shimin.cai81@gmail.com (S.-M. Cai).

https://doi.org/10.1016/j.physa.2018.03.052

correlations. The experimental results in 50 real-world networks also prove the above conclusions.

1. Introduction

Many real-world networks oppose a ubiquitous fat-tailed degree distribution (i.e., scale-free characteristic) [1] and diverse degree correlations [2]. Degree correlations characterize the relationship between the degrees of two connected nodes of the network, and divide the network structure into assortative (positive degree correlation) and disassortative mixing (negative degree correlation) patterns. Epidemic dynamics in complex networks are strongly affected by network topological structure [3–14]. Researchers have found that the network structure, such as the scale-free characteristic and degree correlation has an important influence on the spreading of disease [15–25]. For example, they found the effect of disassortative mixing on epidemic spreading [17,18,26]. That is, the spreading process would slow down if the highly connected nodes (hubs) are more likely to transmit the infection to nodes with low degree [22]. Aim to epidemic dynamics in real-world correlated network, it is of great importance to uncover what affects the outbreak threshold and choose a proper theoretical method to precisely predict outbreak thresholds of these (real-world) correlated networks.

Researchers also have developed many theoretical ways of predicting outbreak thresholds in respect to epidemic spreading models, e.g. susceptible–infected (SI) [23,24], susceptible–infected–susceptible (SIS) [15,16] and susceptible–infected–recovered (SIR) [27]. There are four well-known theoretical methods, each of which applies the topological information differently. The first is the heterogeneous mean-field method (*HMF*) [15,27] that takes into account the degree distribution. The second is the quenched mean-field (*QMF*) method based on the adjacent matrix of the network [28]. The third is the dynamical message passing (*DMP*) method [29] that describes network topology in terms of the non-backtracking matrix. The last is the connectivity matrix (*CM*) method [19] that describes network topology in terms of its connectivity matrix. Because these four methods typically predict different outbreak thresholds for the same (correlated) network, their prediction accuracy is difficult to determine, and how well present theoretical methods predict outbreak thresholds of correlated networks remains unknown.

Recently Wang et al. [30] empirically found that the degree correlations of real world networks affect the outbreak thresholds of SIR spreading dynamics and the prediction accuracy of *HMF*, *QMF* and *DMP*. However, it does not carry out in-depth analysis on that the alterable topological structures of correlated networks relate to the prediction accuracy of outbreak threshold of SIR spreading dynamics. In this paper, we systematically investigate that how predicting outbreak threshold of correlated network is affected by the quantitative changes of degree correlation, degree distribution exponent and network size, and find the fitness of present theoretical methods to predict outbreak threshold of correlated network. Note that we also introduce *CM* to compare with other three present theoretical methods in [30]. Through the extensive experiments in synthetic networks, we show that the outbreak threshold of SIR spreading dynamics reduces with the increasing degree correlation coefficient and network size, but raises with the increasing degree distribution exponent. Furthermore, comparing the simulation results of outbreak thresholds with theoretical ones estimated from these four methods, we find that *CM* and *DMP* have more fitness to precisely predict outbreak thresholds of diverse correlated networks. The above-mentioned results in synthetic networks are proved by empirically findings in 50 real-world networks. Thus, this work provides us to deeply understand the effects of topological structure of correlated networks on precisely predict outbreak threshold of SIR spreading dynamics and find the fitness of the present theoretical methods.

2. Models and methods

2.1. Correlated network model

The correlated network model is able to adjust the degree correlation coefficient r, the degree distribution exponent γ, and the size N. We use the method in [31] to construct such model. Firstly, the uncorrelated configuration model [32] is used to create a network with power-law degree distribution. The degrees of all nodes are limited to a interval $(k_{min}, \sqrt{N})$. Then, we use the biased degree-preserving edge rewiring method to adjust the degree correlation coefficient [33]. The steps are as followed:

(1) Randomly choose two edges and disconnect them.

(2) Then, sort the degrees of the four nodes that associate with the two chosen edges. To generate assortative (dissortaive) networks, the nodes with highest degree should be connected to the node with second highest (lowest) degree, and also connect the remaining two nodes. If one or both of the new edges are already exist in the network, the step should be discarded and randomly choose another two new edges.

(3) Repeat the steps (1) and (2) until the degree correlation coefficient reaches the target value.

According to [2,34], the degree correlation coefficient is defined as

$$r = \frac{\sum_{ij}(A_{ij} - \frac{k_i k_j}{2M})k_i k_j}{\sum_{ij}(k_i\delta_{ij} - \frac{k_i k_j}{2M})k_i k_j}. \quad (1)$$

where A is the adjacency matrix of the network, if node i connects to node j, $A_{ij} = 1$; otherwise, $A_{ij} = 0$, k_i represents the degree of the node i, m is the total number of edges and δ is the Kronecker delta function (which is 1 if $i = j$ and 0 otherwise). The degree correlation is absent in the network when $r = 0$, while $r > 0$ and $r < 0$ indicate positive and negative degree correlation, respectively.

2.2. Epidemic model and theoretical methods

The epidemic spreading dynamics in correlated networks is based SIR model where the nodes are divided into three types, the susceptible, the infected, and the recovered. In SIR model, the infected nodes could transmit the disease to their susceptible neighbors with rate β, and turn to recovery state with rate μ. In this case, there exists a critical value (or the effective epidemic threshold λ_c) of the effective transmission rate $\lambda = \beta/\mu$, above which the final fraction of recovered nodes is finite [5,6].

The four prevalent theoretical methods, *HMF*, *QMF*, *DMP* and *CM* are used to predict the outbreak thresholds of SIR spreading dynamics in correlated networks. In this subsection we will briefly introduce these methods as follows.

HMF predicts the epidemic threshold only using the degree distribution, and can be expressed as

$$\lambda_c^{HMF} = \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}, \tag{2}$$

where $\langle k\rangle$ and $\langle k^2\rangle$ represent the first and second moments of degree distribution $P(k)$ [34], respectively. For a network with power-law degree distribution $P(k) \sim k^{-\gamma}$, the outbreak threshold λ_c^{HMF} predicted by *HMF* is vanishing when $\gamma \leq 3$, otherwise it is finite when $\gamma > 3$ [6]. γ represents the degree distribution exponent of network [3,34]. Although *HMF* is valid for uncorrelated network, it may lost the prediction accuracy for correlated networks because of the degree correlations and dynamic correlations among the states of neighbors [35,36].

QMF, differing from the *HMF*, takes into account the adjacent matrix A of the network [37,38]. It uses the leading eigenvalue Λ_A of the adjacent matrix to predict the outbreak threshold,

$$\lambda_c^{QMF} = \frac{1}{\Lambda_A}. \tag{3}$$

QMF to some extent does not capture the degree correlations and the dynamical correlations among the neighbors and sometimes fails to predict the outbreak threshold of pow-law distributed networks with $\gamma > 3$ [36].

DMP is recently developed to study the SIR model in finite-size networks [29,39,40]. It determines the complete network structure by using the non-backtracking matrix, and thus captures some of dynamical correlations among the states of neighbors. Especially for large sparse networks, the non-backtracking matrix more precisely represents the complete network structure, and it is defined as

$$M = \begin{pmatrix} A & \mathbf{1} - D \\ \mathbf{1} & \mathbf{0} \end{pmatrix}, \tag{4}$$

where $\mathbf{1}$ is a $N \times N$ unit matrix, D is the diagonal matrix with the vertex degrees along its diagonal, and $\mathbf{0}$ is a $N \times N$ null matrix. Thus, *DMP* is able to precisely predict the outbreak threshold via the leading eigenvalue Λ_M of the non-backtracking matrix [41–44], such like

$$\lambda_c^{DMP} = \frac{1}{\Lambda_M}. \tag{5}$$

The last theoretical method, *CM*, considers the degree correlations among the nodes and determines the network structure by using connectivity matrix (CM) $CM_{kk'} = (k-1)P(k'|k)$ [20]. $P(k'|k)$ denotes the probability that a vertex of degree k is pointing to a vertex with degree k'. The outbreak threshold predicted by *CM* is

$$\lambda_c^{CM} = \frac{1}{\Lambda_m}, \tag{6}$$

where Λ_m is the largest eigenvalue of CM. As *CM* takes into account the degree correlations and dynamical correlations among the states of neighbors, it can precisely predict the outbreak threshold of correlated network.

After the brief introductions of four theoretical methods, we simply discuss and compare them when the network is uncorrelated and has a power-law degree distribution and epidemic spreading dynamics is based on SIR model. Firstly, according to [45], *QMF* can predict the outbreak threshold λ_c^{QMF} by the maximum degree k_{max}, $\lambda_c^{QMF} \propto 1/\sqrt{k_{max}}$ when $\gamma > 2.5$, and when $\gamma < 2.5$, $\lambda_c^{QMF} \propto \langle k\rangle/\langle k^2\rangle$. Then, for *CM*, we can easily obtain that $CM_{kk'}$ has an unique eigenvalue $\Lambda_m^{CM'} = (\langle k^2\rangle - \langle k\rangle)/\langle k\rangle$ [20], and $\lambda_c^{CM} = \langle k\rangle/(\langle k^2\rangle - \langle k\rangle)$. And, it is same to λ_c^{HMF} and λ_c^{DMP} [43]. Obviously, *HMF*, *DMP* and *CM* are equal to predict the outbreak threshold when the network is uncorrelated, but the prediction accuracy of these methods is still unknown when the network is correlated.

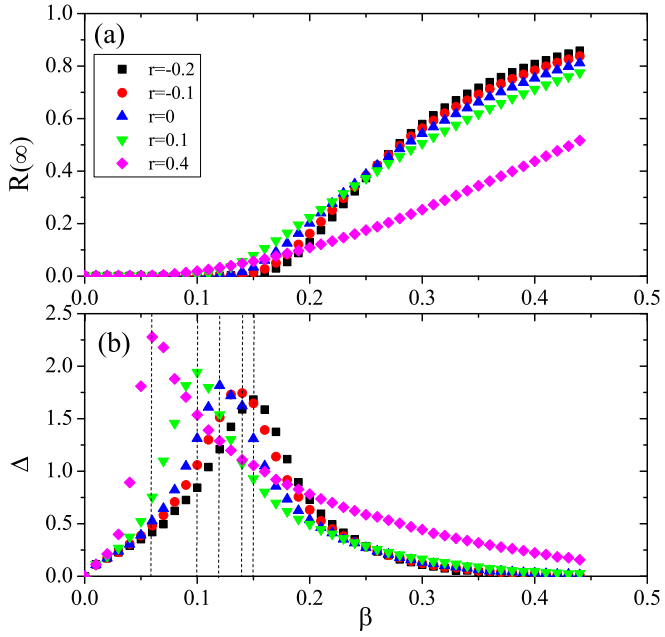

Fig. 1. Overview of SIR spreading dynamics with different degree correlation coefficient r. (a) The outbreak size R_∞ vs. β for $r = 0.4$ (magenta diamonds), 0.1 (green down-triangles), 0 (blue up-triangles), -0.1 (red circles) and -0.2 (black squares). (b) Variability Δ vs. β for $r = 0.4$ (magenta diamonds), 0.1 (green down-triangles), 0 (blue up-triangles), -0.1 (red circles) and -0.2 (black squares). The degree distribution exponent and network size are set as $\gamma = 3.0$ and $N = 10{,}000$, respectively. The results show that the outbreak thresholds reduce with the increasing degree correlations, as well as the final outbreak size.

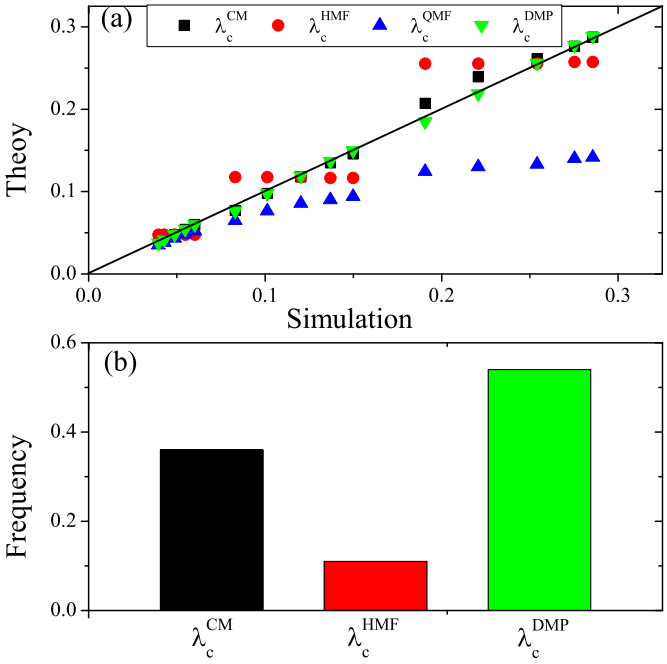

Fig. 2. Compare the prediction accuracy of outbreak thresholds between four theoretical methods and numerical prediction in 15 synthetic networks. The relation of theoretical predictions λ_c^{CM} (black squares), λ_c^{HMF} (red circles), λ_c^{QMF} (blue up-triangles) and λ_c^{DMP} (green down-triangles) versus numerical prediction λ_c. We set the degree distribution exponent $\gamma = 2.2$, 3.0, 4.0, and for each γ, vary the degree correlation coefficient $r = -0.2$, -0.1, 0, 0.1, 0.2.

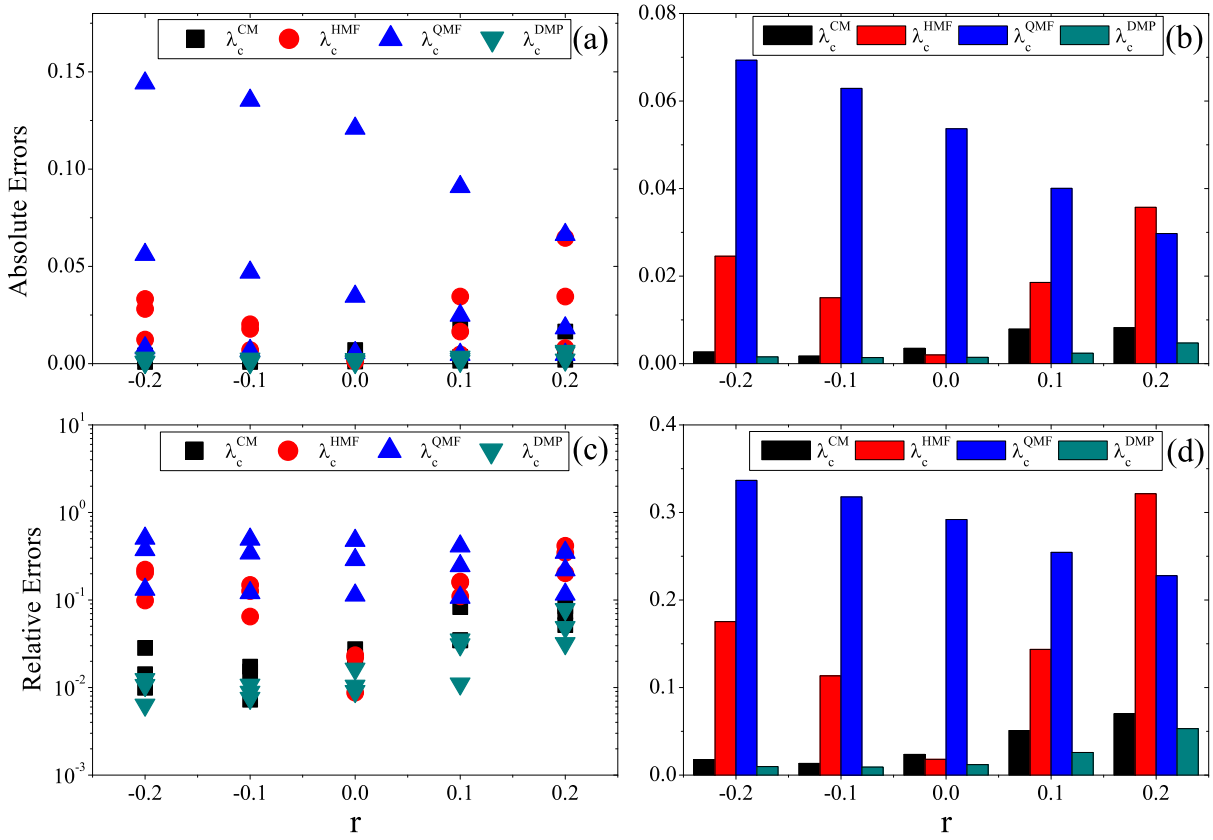

Fig. 3. The influence of degree correlation on the absolute errors and relative errors of four theoretical methods. The first column, (a) and (c) describe the absolute errors and relative errors of theoretical outbreak threshold of four theoretical methods in respect to simulating outbreak threshold. The corresponding average absolute errors (b) and relative errors (d) are shown in the second column. These results are in term of degree correlation coefficient r. We can find that the entirely optimal rank of four theoretical methods according to the average (relative) errors, *DMP* > *CM* > *HMF* > *QMF*. Nevertheless, with the increasing degree correlation coefficients (especial for positive values), *QMF* tends to work better (reduce the average absolute and relative errors) in contrary to other three methods.

3. Simulation results

3.1. Results in synthetic networks

We firstly test the effect of the degree correlations on epidemic threshold and the four theoretical methods, as we known that the degree correlations of the network have a significant effect on the network structure and epidemic threshold [21,22,46]. In the simulations, we set the network size is $N = 10{,}000$, and consider the networks with degree distribution $P(k) \sim k^{-\gamma}$. Also, the minimal and the maximum degrees are set to be $k_{min} = 3$ and $k_{max} = \sqrt{N}$, respectively. To initiate an infection process, we randomly choose five infected nodes as seeds, while the other nodes are in the susceptible state. Without loss of generality, we set the recovery probability $\mu = 1$, and thus the final fraction of recovery nodes $R(\infty)$ and the critical epidemic threshold λ_c is equal to the outbreak size and the outbreak threshold (β_c), respectively. The degree correlations is tuned by using the biased degree-preserving edge rewiring procedure as mentioned in Section 2.1.

The variability Δ is used to numerically determine the epidemic threshold [47–49], which is defined as

$$\Delta = \frac{\sqrt{\langle \rho^2 \rangle - \langle \rho \rangle^2}}{\langle \rho \rangle}, \tag{7}$$

where ρ denotes the epidemic outbreak size $R(\infty)$. To obtain a reliable value of Δ, we repeat at least 3×10^3 independent dynamic realizations on a fixed correlated network and calculate average Δ for each value of unit infection probability β. The peak of Δ correspond to the outbreak threshold β_c of SIR spreading dynamics in correlated network.

Fig. 1 shows that with the network size $N = 10{,}000$ and the degree distribution exponent $\gamma = 3.0$, the outbreak threshold and size of the disease both reduce with the increasing degree correlation coefficients. When the degree correlation coefficient is large (e.g., $r = 0.4$), the nodes with high degree preferably connect to highly connected nodes. Thus, it is easier to form a bigger rich-club structure that makes the spread of disease become more favorable and reduce the outbreak

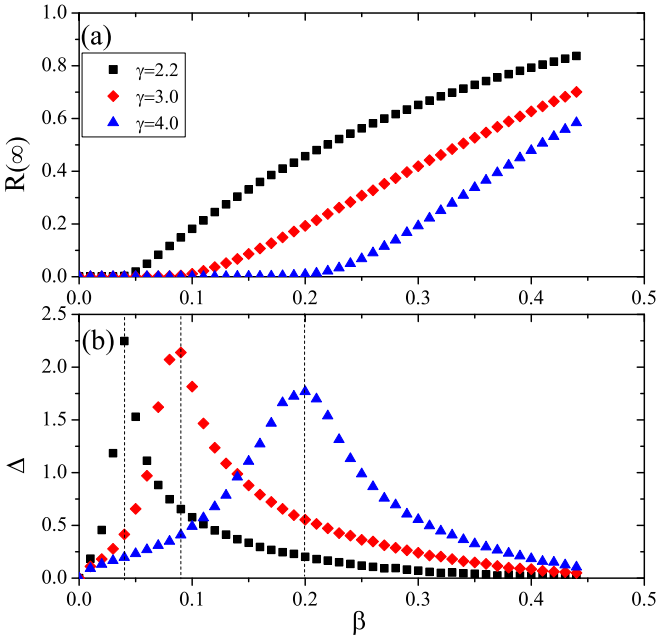

Fig. 4. Overview of SIR spreading dynamics with different degree distribution exponent γ. (a) The outbreak size R_∞ vs. β for $\gamma = 2.2$ (black squares), 3.0 (red diamonds) and 4.0 (blue up-triangles). (b) Variability Δ vs. β for $\gamma = 2.2$ (black squares), 3.0 (red diamonds) and 4.0 (blue up-triangles). The degree correlation coefficient and size of these networks is set $r = 0.2$ and $N = 10{,}000$, respectively. The results show that the outbreak thresholds are increased and the outbreak sizes are suppressed when the network becomes less heterogeneous.

threshold. On the contrary, when the degree correlation coefficient is decreasing (e.g. $r < 0$), highly connected nodes preferably connect to nodes with low degree that makes the network be a star-like structure, which hinders the further spread of disease. On the other hand, with the increasing degree correlation coefficients, the outbreak size $R(\infty)$ becomes smaller at a large β ($\gg \beta_c$), but a litter larger at a small at small β ($> \beta_c$). This result is easily explained as follows: The increasing degree correlation coefficients bring the bigger size of rich-club, and for small β, the existence of the bigger rich-club makes the disease spread more easily, leading to larger $R(\infty)$, while for large β, the disease is more likely to spread between the highly connected nodes and its highly connected neighbors, thus failing to infect a large number of nodes with small degrees and leading to smaller $R(\infty)$.

We further study the influence of degree correlation on the prediction accuracy of the mentioned-above theoretical methods. When the network size is set as $N = 10{,}000$, 15 synthetic networks with power-law degree distribution are constructed by changing degree correlation coefficients $r = -0.2, -0.1, 0, 0.1, 0.2$ and degree distribution exponents $\gamma = 2.2, 3.0, 4.0$. Based on the synthetic networks, we perform the simulations of SIR spreading dynamics and apply four theoretical methods to estimate the corresponding outbreak thresholds. Fig. 2 shows the relation between numerical and theoretical results. Obviously, *DMP* and *CM* predict the outbreak thresholds more precisely, because they both take into account more information of network structure. And, *HMF* only using the degree distribution behaves well in the case of $r = 0$ because it is equal to *DMP* and *CM* (see the introductions and discussions of theoretical methods in Section 2.2). Note that *QMF* only considers the quenching structure of the networks, thus fails to more precisely predict the outbreak threshold of all synthetic networks than other three methods.

In further, in term of degree correlation coefficients, we specially compare the prediction accuracy of four theoretical methods through computing the average absolute error $\Delta(\lambda_c^u) = |\lambda_c^u - \lambda_c|$ and the average relative error $\Delta'(\lambda_c^u) = |\lambda_c^u - \lambda_c|/\lambda_c$ ($u \in \{HMF, QMF, DMP, CM\}$). Fig. 3(b) and (d) show the entirely optimal orders of four theoretical methods according to the average (relative) errors, $DMP > CM > HMF > QMF$. Nevertheless, with the increasing degree correlation coefficients (especial for positive values), *QMF* tends to work better (reduce the average absolute and relative errors) in contrary to other three methods.

As we know that the degree distribution exponents also affect the outbreak thresholds of SIR spreading dynamics in correlated networks, a comparison of four theoretical methods are performed in term of degree distribution exponents. According to the simulations in synthetic networks, we firstly show the outbreak size $R(\infty)$ and the variability Δ as a function of β in Fig. 4. Specifically, for large β, $R(\infty)$ reduces and λ_c increases with the increasing γ. This result can be qualitatively explained that when γ increases, the networks become less heterogeneous and thus lack of highly connected nodes that suppress the outbreak of disease spreading and weaken $R(\infty)$ for large β.

In further, in term of degree distribution exponent γ, we show in Fig. 5 the (average) absolute and relative errors of λ_c^{HMF}, λ_c^{QMF}, λ_c^{DMP}, and λ_c^{CM} in respect to λ_c. Specifically, in Fig. 5(b) and (d), we can see that when the network become less

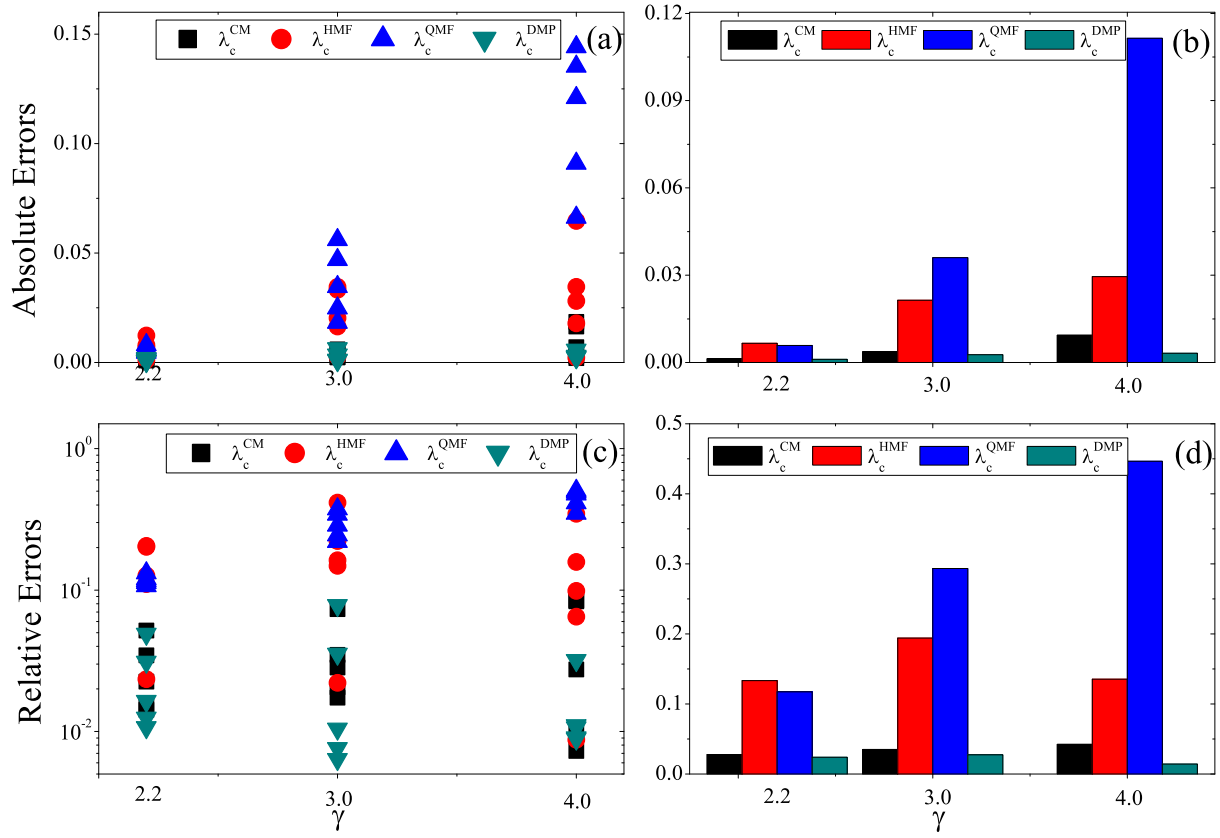

Fig. 5. The influence of degree distribution exponents on the absolute errors and relative errors of the four theoretical methods. In the first column, (a) and (c) are the absolute errors and relative errors of theoretical outbreak thresholds of four theoretical methods in respect to simulating outbreak thresholds. The corresponding average absolute errors (b) and relative errors (d) are shown in the second column. These results are in term of degree distribution exponents, and confirm that the optimal rank of four theoretical methods is *DMP* > *CM* > *HMF* > *QMF*. Furthermore, *QMF* are very sensitive to larger degree distribution exponent.

heterogeneous (i.e., larger γ) λ_c^{DMP} robustly approaches to λ_c in comparison of other three methods, and for *HMF* and *QMF* the (average) absolute and relative errors obviously increase. These results confirm that *DMP* is most precise for predicting epidemic threshold of SIR spreading dynamics in correlated networks with power-law degree distribution.

Finally, we also investigate the influence of network size on the prediction accuracy of SIR spreading dynamics in correlated networks. Through presetting $r = 0.1$ and $\gamma = 2.2$, we increase the sizes of correlated networks with power-law degree distribution from 1000 to 16,000 (the increment step is 2^n). Based on the SIR spreading dynamics in these synthetic networks, we perform the simulations and the theoretical analysis on prediction accuracy of outbreak thresholds. Fig. 6 shows that the outbreak thresholds reduce with the increasing network size and the outbreak sizes $R(\infty)$ for large β are almost independent of the network size. According to network model in Section 2.1, the large-degree (or highly connected) nodes increases with the network size, which promotes the spreading of disease.

Like above-mentioned analysis of degree correlation and degree distribution exponent, we also check the influence of network size on the prediction accuracy of four theoretical methods by 5 positively ($r = 0.1$) and 5 negatively ($r = 0.1$) correlated networks with power-law degree distribution $\gamma = 2.2$. Fig. 7 shows the relations between theoretical and numerical predictions, which suggests that λ_c^{DMP} and λ_c^{CM} are more precisely close to the outbreak thresholds of numerical simulations. Thus, *DMP* and *CM* are also robust to the changing network size, as well as the degree correlation and degree distribution exponent.

Then, in term of network size, we shown the (average) absolute and relative errors of λ_c^{HMF}, λ_c^{QMF}, λ_c^{DMP} and λ_c^{CM} in respect to λ_c (see Fig. 8). Specifically, in Fig. 8(b) and (d), we can see that with the increasing network size, the average absolute and the relative errors of λ_c^{DMP} and λ_c^{CM} are much smaller than *HMF* and *QMF* and robust to network size, which confirms the better predictions of *DMP* and *CM*.

3.2. Results in real-world networks

We provide a comparison of four theoretical methods for predicting outbreak threshold of SIR spreading dynamics in synthetic networks, and confirm that *DMP* and *CM* show much better theoretical predication accuracy in respect to

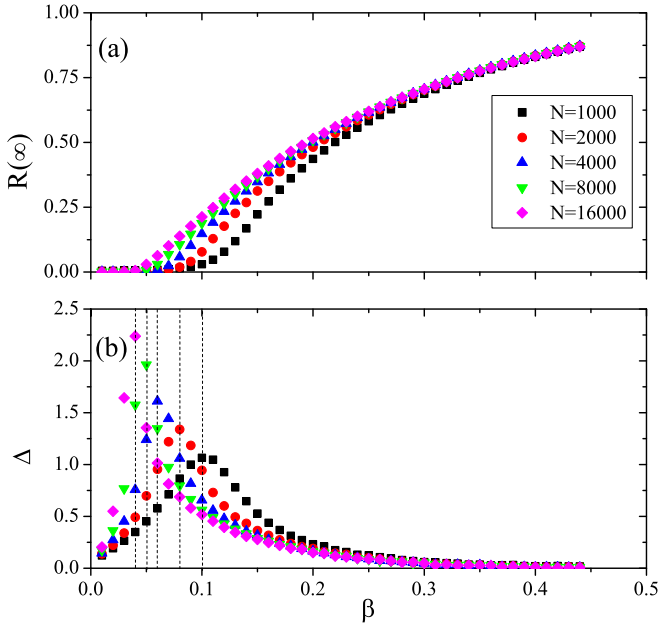

Fig. 6. Overview of SIR spreading dynamics with different network size N. (a) Final outbreak size R_∞ vs. β for $N = 1000$ (black squares), 2000 (red circles), 4000 (blue up-triangles), 8000 (green down-triangles) and 16 000 (magenta diamonds). (b) Variability Δ vs. β for $N = 1000$ (black squares), 2000 (red circles), 4000 (blue up-triangles), 8000 (green down-triangles) and 16 000 (magenta diamonds). We fix the degree correlation coefficient $r = 0.1$ and the degree distribution exponent $\gamma = 2.2$. We can see that the outbreak thresholds reduces with the increasing network size and the outbreak sizes $R(\infty)$ for large β are almost independent of the network size.

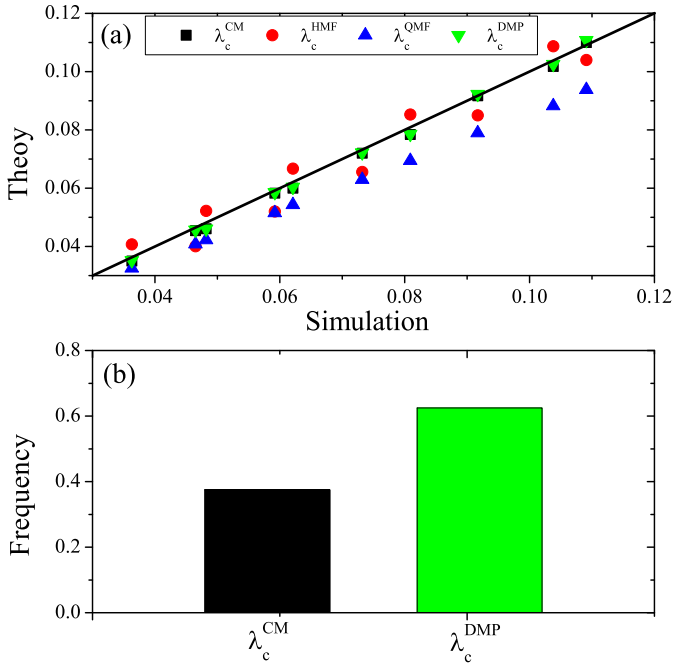

Fig. 7. Compare the prediction accuracy of outbreak thresholds between four theoretical methods and numerical prediction in 10 synthetic networks. Theoretical predictions of λ_c^{CM} (black squares), λ_c^{HMF} (red circles), λ_c^{QMF} (blue up-triangles) and λ_c^{DMP} (green down-triangles) versus numerical prediction λ_c. For positively ($r = 0.1$) and negatively ($r = -0.1$) correlated networks, we set the degree distribution exponent $\gamma = 2.2$ and the network size $N = 1000, 2000, 4000, 8000, 16\,000$.

numerical simulations. In this section, we extend the comparison of four theoretical methods via 50 real world networks. The topological structures of 50 real world networks and their theoretical outbreak thresholds of SIR spreading dynamics are

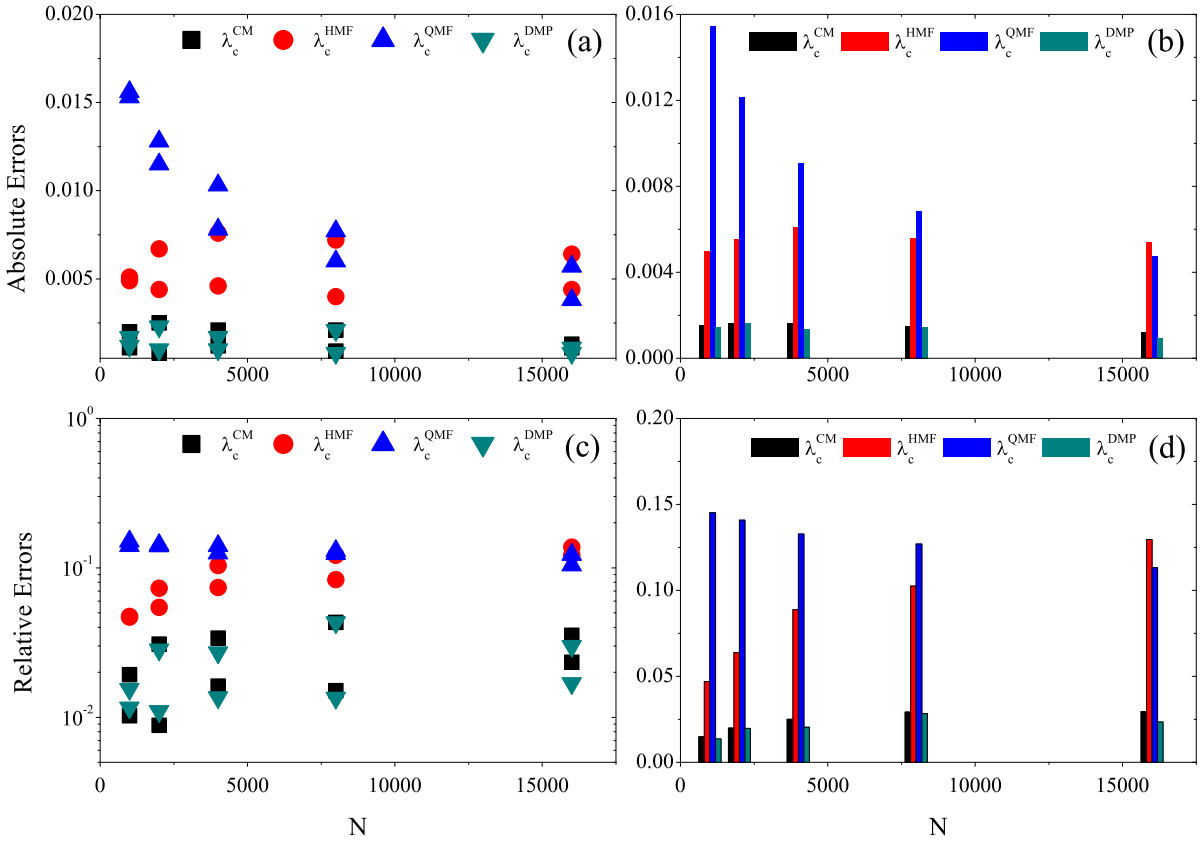

Fig. 8. The influence of network size on the absolute errors and relative errors of four theoretical methods. In the first column, (a) and (c) are the absolute errors and relative errors of theoretical methods in term of network size N, and the corresponding average absolute errors (b) and relative errors (d) are shown in the second column. We can find that the prediction accuracy of *DMP* and *CM* are significantly better than *HMF* and *QMF*.

present (see supporting material). Note that 50 real world networks do not strictly obey the power-law distribution, so that the degree distribution exponents cannot precisely estimated. And, we mainly analyze the influence of degree correlations on the prediction accuracy of outbreak thresholds of SIR spreading dynamics in 50 real world networks.

Firstly, we explore the relations between the theoretical and numerical outbreak thresholds of SIR spreading dynamics in 50 real world networks. Fig. 9 shows the prediction accuracy of four theoretical methods in respect to numerical simulations, which suggests that *DMP*, *CM* and *QMF* behave relatively better, and *HMF* obviously works in the worst.

Then, to more precisely evaluate the performance of four theoretical methods, we statistically divide 50 real world networks in term of degree correlation. The interval of degree correlation coefficients of 50 real world networks is $[-0.878, 0.267]$, of which the bin $(r - \Delta r/2, r + \Delta r/2)$ is set with the size $\Delta r = 0.1$. For each bin, compute the (average) absolute errors and relative errors of λ_c^{HMF}, λ_c^{QMF}, λ_c^{DMP} and λ_c^{CM} in respect to λ_c in respect to λ_c (see Fig. 10). We can see that in most cases, *CM* perform better than other three methods. However, when the degree correlation coefficient is too small, *DMP* is the best prediction while *HMF* works in the worst. Thus, we claim that the *CM* and *DMP* are more proper to predict epidemic threshold of SIR spreading dynamics in correlated networks.

4. Conclusion

We have systematically investigated the influence of degree correlation, degree distribution exponent and network size on the epidemic thresholds of SIR spreading dynamics in correlated networks and provided a comparison of present four theoretical methods to precisely predict the epidemic thresholds of correlated networks. Firstly, through presetting the network size and the degree distribution exponent, we explore that how the changing degree correlations affect the epidemic thresholds of SIR spreading dynamics. The simulation results suggest that the gradually increasing degree correlation coefficients reduce the outbreak threshold and suppress the outbreak size because more nodes form the rich-club structure promoting the spreading of disease. And, the comparison of four theoretical methods in respect to numerical simulations indicates that *DMP* performs the best predictions of outbreak threshold of synthetically correlated networks, and a little better than *CM*.

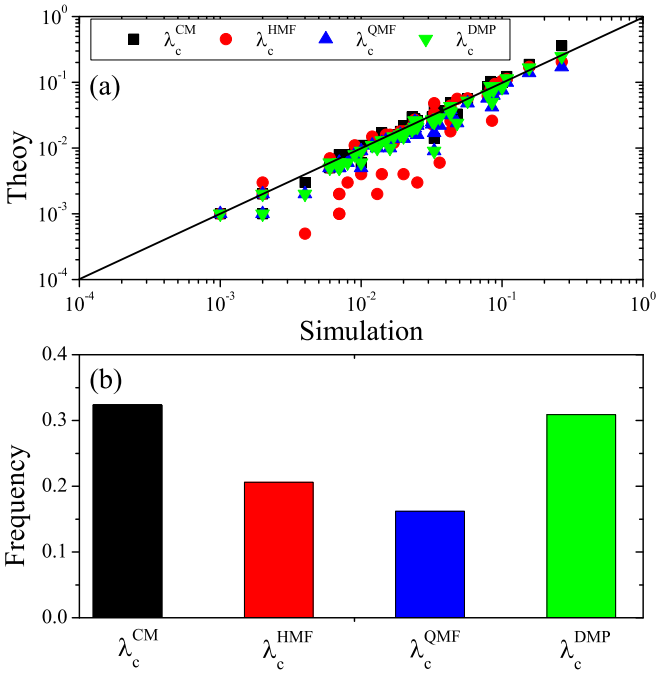

Fig. 9. Compare the prediction accuracy of outbreak thresholds between four theoretical methods and numerical prediction in 50 real-world networks. Theoretical predictions of λ_c^{CM} (black squares), λ_c^{HMF} (red circles), λ_c^{QMF} (blue up-triangles) and λ_c^{DMP} (green down-triangles) versus numerical prediction λ_c. We can see that *DMP*, *CM* and *QMF* behave relatively better, and *HMF* works in the worst.

Then, we illustrate the influence of degree distribution exponent on the epidemic threshold and the prediction accuracy of outbreak threshold via four theoretical methods. With the increasing degree distribution exponent, the network becomes less heterogeneous, which makes nodes with high degree decrease. Thus, it results that the spreading of disease becomes difficult, which increases the outbreak threshold but suppresses the outbreak size. And, in term of degree contribution exponent, the comparison of four theoretical methods in respect to numerical simulations represents the similar conclusion found in term of degree correlation.

In addition, we also examine the influence of network size on the epidemic threshold and the prediction accuracy of outbreak threshold via four theoretical methods. We find that the correlated networks with larger size easily promote the epidemic outbreak, however, little affect the outbreak size. The reason may be that with the increase of the network size, the nodes with higher degree will increase according to network model, which speeds up the spreading process. And, in term of network size, *DMP* and *CM* behave approximately same performance.

Finally, as the comparison of four theoretical methods is based on the analysis and numerical simulations performed in synthetically (correlated) networks, we extend it based on 50 real world (correlated) networks. In term of degree correlation, 50 real world networks are statistically classified into several bins. For each bin, we compare the prediction accuracy of outbreak thresholds via four theoretical methods in respect to numerical simulations. Overall, *CM* performs the best in most of cases, however, *DMP* behaves better when the degree correlation coefficient is too small. Thus, we confirm that *CM* and *DMP* are more proper to predict epidemic threshold of SIR spreading dynamics in real-world correlated networks.

Precisely predicting epidemic thresholds in correlated networks is profoundly significant in the field of spreading dynamics. Our results suggest that the epidemic threshold of the network is not only related to the degree correlation, but also to the degree distribution and the size of the network. The findings expand our understanding of epidemic thresholds and provide ways of determine the fitness of prevalent theoretical method to predict the epidemic threshold in a variety of correlated networks. It is worth noting that in synthetically correlated networks *DMP* method robustly performs in the best, but in the real-world networks, it only works better when degree correlation coefficient is very small. This case may to some extent relate to other factors of network structure, such as the diverse heterogeneity and size of real world networks, which is still a question worthy of further discussion.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61673086, 11575041, 71601030 and 61673085) and the Fundamental Research Funds for the Central Universities, China (Grant Nos. ZYGX2015J153 and ZYGX2016J058). The Center for Polymer Studies of Boston University is supported by NSF, United States Grants PHY-1505000, CMMI-1125290, and CHE-1213217, by DTRA, United States Grant HDTRA1-14-1-0017, and by DOE, United States Contract DE-AC07-05Id14517.

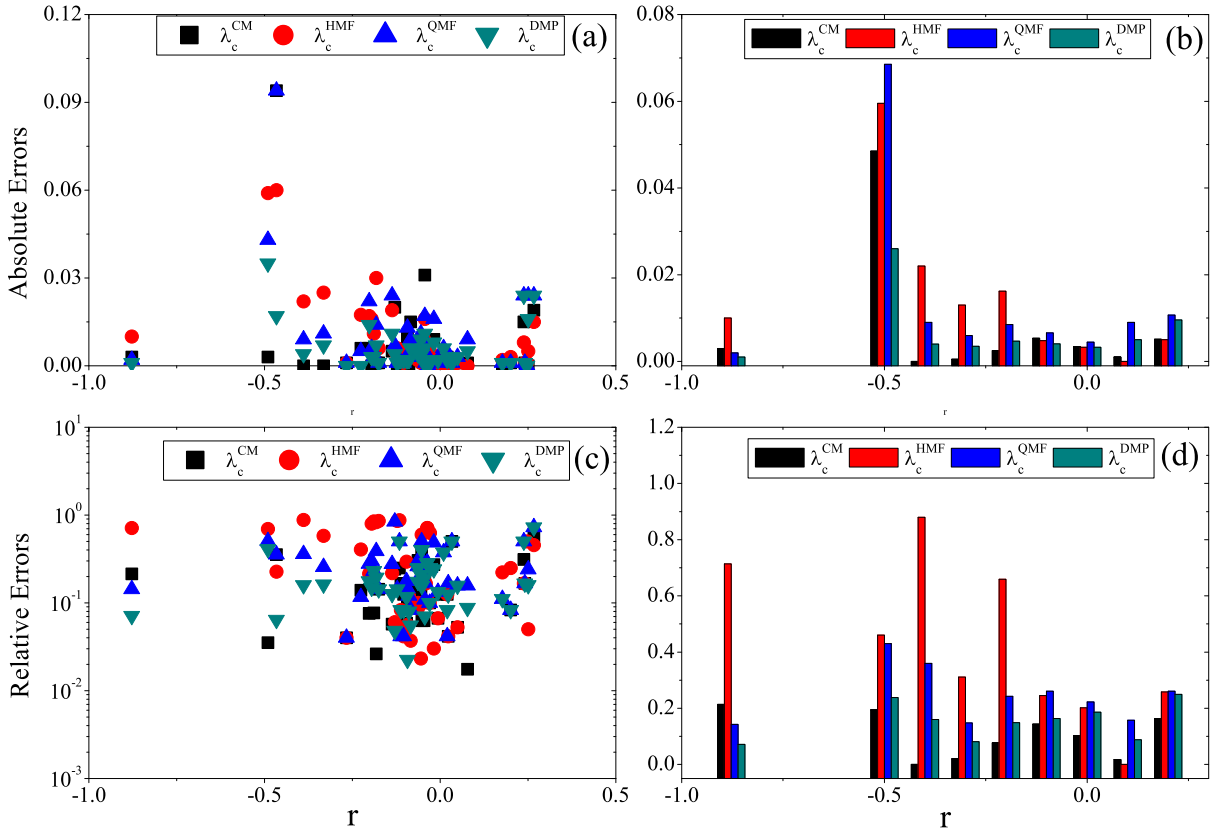

Fig. 10. The influence of degree correlation on the absolute errors and relative errors of four theoretical methods. In the first column, (a) and (c) are the absolute errors and relative errors of theoretical methods in respect to simulating outbreak thresholds. The corresponding average absolute errors (b) and relative errors (d) are shown in the second column. The results are in term of degree correlation, and confirm that in most of networks.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physa.2018.03.052.

References

[1] A.-L. Barabási, R. Albert, Science 286 (5439) (1999) 509–512.
[2] M.E. Newman, Phys. Rev. Lett. 89 (20) (2002) 208701.
[3] R. Albert, A.-L. Barabási, Rev. Modern Phys. 74 (1) (2002) 47.
[4] M.E.J. Newman, SIAM Rev. 45 (2) (2003) 167–256.
[5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Phys. Rep. 424 (4) (2006) 175–308.
[6] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Rev. Modern Phys. 87 (3) (2015) 925.
[7] H. Liao, A. Zeng, Sci. Rep. 5 (2015) 11404.
[8] J.-Q. Kan, H.-F. Zhang, Commun. Nonlinear Sci. Numer. Simul. 44 (2017) 193–203.
[9] H.-F. Zhang, P.-P. Shu, Z. Wang, M. Tang, M. Small, Appl. Math. Comput. 294 (2017) 332–342.
[10] H. Liao, M.S. Mariani, M. Medo, Y.-C. Zhang, M.-Y. Zhou, Phys. Rep. 689 (2017) 1–54.
[11] W. Wang, M. Tang, S.H. Eugene, L.A. Braunstein, Rep. Progr. Phys. 80 (3) (2017) 036603.
[12] H.-X. Yang, M. Tang, Y.-C. Lai, Phys. Rev. E 91 (6) (2015) 062817.
[13] C. Liu, X.-X. Zhan, Z.-K. Zhang, G.-Q. Sun, P.M. Hui, New J. Phys. 17 (11) (2015) 113045.
[14] C. Liu, Z.-K. Zhang, Commun. Nonlinear Sci. Numer. Simul. 19 (4) (2014) 896–904.
[15] R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. 86 (14) (2001) 3200.
[16] R. Pastor-Satorras, A. Vespignani, Phys. Rev. E 63 (6) (2001) 066117.
[17] V.M. Eguiluz, K. Klemm, Phys. Rev. Lett. 89 (10) (2002) 108701.
[18] D. Volchenkov, L. Volchenkova, P. Blanchard, Phys. Rev. E 66 (4) (2002) 046137.
[19] M. Boguná, R. Pastor-Satorras, Phys. Rev. E 66 (4) (2002) 047104.
[20] M. Boguná, R. Pastor-Satorras, A. Vespignani, arXiv preprint Cond-Mat/0301149 (2003).
[21] Y. Moreno, J.B. Gómez, A.F. Pacheco, Phys. Rev. E 68 (3) (2003) 035103.
[22] M. Boguná, R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. 90 (2) (2003) 028701.
[23] M. Barthlemy, A. Barrat, R. Pastorsatorras, A. Vespignani, Phys. Rev. Lett. 92 (17) (2004) 178701.
[24] T. Zhou, J.-G. Liu, W.-J. Bai, G. Chen, B.-H. Wang, Phys. Rev. E 74 (5) (2006) 056109.

[25] Z.-K. Zhang, C. Liu, X.-X. Zhan, X. Lu, C.-X. Zhang, Y.-C. Zhang, Phys. Rep. 651 (2016) 1–34.
[26] C.P. Warren, L.M. Sander, I.M. Sokolov, Phys. Rev. E 66 (5) (2002) 056105.
[27] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Eur. Phys. J. B 26 (4) (2002) 521–529.
[28] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, Y. Moreno, Europhys. Lett. 89 (3) (2010) 38009.
[29] B. Karrer, M.E. Newman, Phys. Rev. E 82 (1) (2010) 016101.
[30] W. Wang, Q.-H. Liu, L.-F. Zhong, M. Tang, H. Gao, H.E. Stanley, Sci. Rep. 6 (2016).
[31] L. Gao, W. Wang, L. Pan, M. Tang, H.F. Zhang, Sci. Rep. (2016).
[32] M. Catanzaro, M. Boguná, R. Pastor-Satorras, Phys. Rev. E 71 (2) (2005) 027103.
[33] R. Xulvi-Brunet, I.M. Sokolov, Phys. Rev. E 70 (6) (2004) 066102.
[34] M. Newman, Networks: An Introduction, Oxford University Press, 2010.
[35] S.C. Ferreira, C. Castellano, R. Pastor-Satorras, Phys. Rev. E 86 (4) (2012) 041125.
[36] C. Castellano, R. Pastor-Satorras, Phys. Rev. Lett. 105 (21) (2010) 218701.
[37] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, C. Faloutsos, ACM Trans. Inf. Syst. Secur. (TISSEC) 10 (4) (2008) 1.
[38] P. Van Mieghem, J. Omic, R. Kooij, IEEE/ACM Trans. Netw. 17 (1) (2009) 1–14.
[39] M. Shrestha, S.V. Scarpino, C. Moore, Phys. Rev. E 92 (2) (2015) 022821.
[40] A.Y. Lokhov, M. Mézard, L. Zdeborová, Phys. Rev. E 91 (1) (2015) 012811.
[41] F. Radicchi, Phys. Rev. E 91 (1) (2015) 010801.
[42] T. Martin, X. Zhang, M. Newman, Phys. Rev. E 90 (5) (2014) 052808.
[43] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, P. Zhang, Proc. Natl. Acad. Sci. 110 (52) (2013) 20935–20940.
[44] B. Karrer, M.E. Newman, L. Zdeborová, Phys. Rev. Lett. 113 (20) (2014) 208702.
[45] F. Chung, L. Lu, V. Vu, Proc. Natl. Acad. Sci. 100 (11) (2003) 6313–6318.
[46] Z. Rong, Z.-X. Wu, Europhys. Lett. 87 (3) (2009) 30001.
[47] P. Crepey, F.P. Alvarez, M. Barthélemy, Phys. Rev. E 73 (4) (2006) 046131.
[48] P. Shu, M. Tang, K. Gong, Y. Liu, Chaos 22 (4) (2012) 043124.
[49] P. Shu, W. Wang, M. Tang, Y. Do, Chaos 25 (6) (2015) 063104.